# Random Access for Machine-Type Communications

*Israel Leyva-Mayorga[1], Cedomir Stefanovic[1], Petar Popovski[1], Vicent Pla[2], and Jorge Martinez-Bauset[2]*

[1] *Aalborg University, Aalborg, Denmark*
[2] *Universitat Politècnica de València, Valencia, Spain*

## Introduction

The 3rd Generation Partnership Project (3GPP) has recently concluded the first phase of 5G standardization. 5G promises versatile support of heterogenous services with different requirements in terms of data rates, latency, and/or number of connected devices. The success of 5G crucially depends on its ability to solve the shortcomings that were identified in 4G, one of which is the efficient support of machine-type communications (MTC).

MTC, also known as machine-to-machine (M2M) communication, stands for the autonomous data exchange between devices and enables a wide range of applications (3GPP TS 22.368 V14.0.1 2017), like smart metering, fleet management, e-health care, asset tracking, etc. Commonly, MTC applications incorporate a large number of wireless devices with low communication and processing capabilities that perform simple tasks, such as sporadic data collection and transmission. These data are typically small, pertaining to short instructions, reports on the state of the system components, environmental data, etc. Therefore, the design of MTC devices is mainly focused on reducing size, cost, and power consumption.

The characteristics of MTC applications greatly differ from those of human-to-human (H2H) applications, specifically in terms of traffic characteristics. In comparison to H2H traffic, MTC traffic exhibits a much lower amount of mobility and of transmitted data while potentially involving a much greater degree of spatial and temporal correlation. That is, MTC data generation process can be highly synchronized (3GPP TS 22.368 V14.0.1 2017). In addition, MTC devices switch back and forth from connected to disconnected mode to save power, which leads to performing frequent access requests to the network.

Due to the success of MTC applications, MTC connections will grow from roughly $6 \times 10^9$ in 2017 to nearly $15 \times 10^9$ by 2022 (Cisco 2017). Hence, applications with tens of thousands or more MTC devices, known as massive machine-type communications (mMTC) applications, will become widespread in the near future. These applications are particularly difficult to support due to the large number of simultaneous access requests,

*Wiley 5G Ref.* Edited by Rahim Tafazolli, Chin-Liang Wang and Periklis Chatzimisios.

DOI: 10.1002/9781119471509.w5GRef031

which can easily exceed the signaling capabilities of the access network. This problem exists even if the access requests are not synchronized, but synchronization among them greatly aggravates it.

Ultrareliable low latency communications (URLLC) pertain to a different set of MTC applications, involving a much lower number of communicating devices compared to mMTC, but with more stringent latency and reliability requirements, e.g. latency of a few milliseconds and a packet error probability of $10^{-5}$.

A major challenge of 5G is to guarantee the widely different requirements of mMTC and URLLC. Furthermore, cellular networks up to 4G were designed for H2H communication and numerous features of 4G were simply adopted in the first phase of standardization of 5G with minor to no modifications. In this respect, a feature that is particularly ill-suited to handle MTC traffic is the method used to grant the initial access to the cellular network: the random access (RA) procedure.

The RA procedure consists of a four-message exchange: preamble, grant, connection request, and contention resolution. These must be performed between the cellular base station (BS) and every wireless device, i.e. *user equipment (UE),* that needs to switch from idle to connected mode. In other words, the RA procedure is a *grant-based protocol*, starting with the preamble (metadata) sent by a UE and data is transmitted only after the four-message exchange is successfully completed, as described in the section titled "5G Random Access (RA) Procedure".

Several studies have found that the initial access attempts that occur in mMTC applications may severely overload the access network (Laya et al. 2014; Osti et al. 2014; Tello-Oquendo et al. 2018, p. 3512). Moreover, the excessive signaling during the RA procedure highly compromises the latency constrains of URLLC MTC; the transmission and processing of the four messages that comprise the RA procedure is in the order of 15 ms (3GPP TS 36.912 V15.0.0 2018, Table 16.2.1-1; Leyva-Mayorga et al. 2017, p. 7793).

Standardization efforts to support mMTC in LTE-A led to the development of the narrowband Internet of Things (NB-IoT) and the LTE-M specifications. Both of these were completed in Release 13 of the 3GPP specifications and focus on providing great power efficiency, low bandwidth utilization, and enhanced coverage at a reduced hardware cost. The NB-IoT devices are expected to remain active for up to 15 years without the need of battery replacements and to communicate at a distance of up to ten kilometers from the BSs (3rd Generation Partnership Project 2017). Nevertheless, the RA procedure in NB-IoT and LTE-M is similar to that of 4G and 5G with only minor differences in the physical layer. Hence, the problems of the cellular RA procedure persist in NB-IoT and LTE-M.

Access control mechanisms, in the form of barring schemes, have been included in the 3GPP technical specifications (3GPP TS 22.011 V16.4.0 2018, Section 4; 3GPP TS 36.331 V15.0.0 2018). Barring schemes aim to solve congestion problems by broadcasting a set of barring parameters to prevent the access requests from certain subsets of UEs. In principle, the barring parameters must be selected according to the number of UEs that attempt to access the BS. In practice, it is rather complicated to set adequate values for the barring parameters and the 3GPP standards do not provide guidance on this matter. In contrast, the research literature on adequate selection of barring parameters is vast (De Andrade et al. 2017; Duan et al. 2016; Leyva-Mayorga et al. 2019; Lin et al. 2014; Tavana et al. 2018; Wang and Wong 2015).

Further approaches to support mMTC in 5G include cooperative RA and other enhancements to the RA procedure. However, these may not be sufficient to guarantee the latency requirements of URLLC MTC. An appealing solution to support both mMTC and URLLC MTC is to replace or complement the current grant-based random access (GBRA) procedure with a *grant-free random access (GFRA) protocol*. In GFRA, UEs contend with their data packets in a RA fashion, rather than contending with access requests for the purpose of obtaining a grant to transmit the data packets. GFRA protocols are widely used in non-3GPP IoT solutions such as LoRaWAN (LoRa Alliance Technical Committee 2017) and Sigfox (www.sigfox.com). In summary, further research is needed to identify the preferred RA and/or access control mechanisms to efficiently handle mMTC and URLLC MTC in 5G.

The rest of the article is organized as follows. A general view of communication models for MTC is presented in the section titled "Communication Models for MTC", providing the theoretical background to assess the performance of the RA procedure. The RA procedure, as defined for 5G and legacy 4G, along with the changes introduced in NB-IoT and LTE-M, is described in detail in the section titled "5G Random Access (RA) Procedure". The limitations of the RA procedure under mMTC applications and the 3GPP access control mechanisms are described in the section titled "Random Access in mMTC Applications". A survey of proposed improvements to access control mechanisms included in the standards is presented in the section titled "Improvements to Existing Protocols". Finally, a summary of emerging technologies and the conclusions are presented in the section titled "Emerging Technologies and Conclusions".

## Communication Models for MTC

In this section, we first introduce a general channel model for the RA in MTC applications, from which the channel models for GBRA and GFRA are subsequently derived.

Consider a wireless network with a star topology, where the nodes communicate directly with the BS. The network operates in a slotted channel: i.e. the time is divided into slots and the whole duration of the slot is used for transmission. Next, let $N$ be the total number of users within the area of coverage of the BS. The general channel model of a single quasi-static channel that remains constant during a slot can be formulated as

$$y = \sum_{n=1}^{N} h_n a_n x_n + z + w \tag{1}$$

where, for the $n$th user at the given time slot,

- $y$ is the received signal.
- $h_n$ is the wireless channel coefficient.
- $a_n$ is an indicator variable; $a_n = 1$ if the user is active and $a_n = 0$ otherwise.
- $x_n$ is the transmitted signal.
- $z$ is the noise.
- $w$ is the interference.

In other words, Eq. (1) is the model of a block-fading channel whose coefficients remain constant throughout the transmission of $x_n$.

It is in the very nature of RA protocols that the value of $a_n$ at a given slot is not known beforehand. Therefore, the main task of the BS is to determine the values of $a_n$ for the $N$ users from the received signal $y$ at each slot and, if $a_n = 1$, decode $x_n$. This task is complicated due to numerous factors. In particular, the channel coefficient $h_n$, the noise $z$, and the interference $w$ may not be known. These, together with the unknown values of $a_n$, contribute to the overall uncertainty, which also includes the uncertainty about the desired signals $\{x_n\}$.

A simplified model can be obtained from (1) under the following conditions:

1) The network operates in licensed spectrum, so the interference $w$ is under control and can be neglected.
2) The noise power is sufficiently low and can be sufficiently low and can be neglected.
3) Wireless devices can obtain accurate channel-state information (CSI) and, if sufficiently strong, the channel can be inverted such that $h_n = 1$. Otherwise, if the channel is too weak to be inverted, we set $a_n = 0$ for that device. This is equivalent to setting $h_n = 1$ for all $n$ in (1) and work with the uncertainty in $a_n$ only.

Under these conditions, the general channel model becomes

$$y = \sum_{n=1}^{N} a_n x_n \tag{2}$$

This simplified model will be used throughout the article, unless otherwise stated.

The RA protocol defined for 5G is grant based. Therefore, the focus of the remainder of this section is on GBRA protocols, while GFRA is only briefly described at the end of this section.

### Grant-Based Random Access (GBRA)

GBRA comprises the transmission of an access request by a user, which contains some sort of the user ID, followed by a grant from the BS. Only the users that receive a grant from the BS can proceed to data transmission. Finally, an acknowledgment is sent if data is correctly received. Otherwise, a retransmission may be requested.

To model a general GBRA protocol, we define the events $\mathcal{S}_\mathrm{r}$ and $\mathcal{S}_\mathrm{g}$ as the successful reception of an access request and access grant. The probability of a successful access in GBRA is given as

$$\Pr[\mathcal{S}_{\mathrm{GBRA}}] = \Pr[\mathcal{S}_\mathrm{r} \text{ and } \mathcal{S}_\mathrm{g}] = \Pr[\mathcal{S}_\mathrm{r}] \Pr[\mathcal{S}_\mathrm{g}] \tag{3}$$

$\Pr[\mathcal{S}_\mathrm{g}]$ depends not only on the conditions of the wireless medium but also on the available resources for signaling at the BS, since the number of access grants may be limited.

In the simplest GBRA model, the access requests follow a slotted ALOHA protocol. A collision occurs if two or more users transmit their requests simultaneously. In the basic collision channel model, the BS is unable to decode the any of the collided requests and, as a result, no grant is transmitted in response. In other words, the access request is successfully received only if the number of access requests in the slot is $A = \sum_{n=1}^{N} a_n = 1$. Afterward, the grant, data, and acknowledgment are transmitted through dedicated resources.

The slotted ALOHA protocol is the basis for numerous RA protocols. An immediate extension is the multichannel slotted ALOHA, where the channels are realized by employing orthogonal resources, such as codes or frequencies, and multiple users can access the BS using different resources simultaneously (i.e. see Orthogonal Multiple Access).

As described later in detail, the first step of the 5G RA procedure, preamble transmission, is equivalent to the multichannel allotted ALOHA, where the number of channels equals the number of available preambles. To model the multichannel slotted ALOHA, let $\mathcal{X} = \{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_K\}$ be the set of $K$ orthogonal resources available in a slot, selected uniformly at random by the active users. Denote by $i_n \in \{1, 2, \ldots, K\}$ the index of the resource selected by user $n$, and denote by $\boldsymbol{x}_n = [x_{n1}, x_{n2}, \ldots, x_{nK}]^T$ its transmitted signal, where

$$x_{nk} = \begin{cases} \mathcal{X}_k, & \text{if } a_n > 0 \text{ and } i_n = k \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

Building on this, the simplified model for multichannel slotted ALOHA is given as

$$\boldsymbol{y} = \sum_{n=1}^{N} \boldsymbol{x}_n \tag{5}$$

where $\boldsymbol{y} = [y_1, y_2, \ldots, y_K]^T$. The analysis of the capacity of multichannel slotted ALOHA GBRA is presented in the section titled "Random Access in mMTC Applications".

### Grant-Free Random Access (GFRA)

The basic idea of GFRA is to skip the reservation phase and directly transmit the data along with the metadata required for UE identification. GFRA protocols are common in non-3GPP IoT solutions such as LoRaWAN and Sigfox, as well as in traditional wireless sensor networks.

Skipping the access request and grant transmissions greatly reduces the complexity of the RA and may also reduce the access latency and increase the success probability when compared to GBRA. In particular, GFRA can outperform GBRA if the following conditions are met:

1) The same number of orthogonal resources are used.
2) The size of the data packets relative to the size of control packets is small.
3) Data packets are transmitted sporadically.

Obviously, mMTC and URLLC MTC applications may greatly benefit from GFRA, as they are characterized by the infrequent transmission of short data. On the other hand, GBRA is preferred in applications with frequent data transmissions, where bursts of data packets are transmitted immediately after a single access request and/or the size of the data packets is large. Specifically, if multiple packets are to be transmitted, the GBRA is performed only once, and the probability of success of the data packet transmissions is close to one, as they happen in reserved resources. In addition, if the data packet is large, the amount of wasted resources (energy and bandwidth) when losing a data packet in GFRA will be greater than losing an access request in GBRA.

## 5G Random Access (RA) Procedure

This section describes the two modes for the RA procedure in 5G: contention-based random access (CBRA) and contention-free random access (CFRA), as defined in its first phase of 5G standardization (3GPP TS 38.300 V15.3.1 2018; 3GPP TS 36.321 V15.2.0 2018, Section 5.1). Both modes correspond to a GBRA protocol that operates in a slotted channel, such as the one defined by (5). In 3GPP cellular networks, subframes are a minimum unit for the scheduling of RA requests in the time domain; a radio frame is a set of 10 consecutive subframes beginning at subframe 0.

Before initiating the RA procedure, the UEs must acquire the basic network configuration parameters and the Physical Random Access Channel (PRACH) parameters; the latter correspond to the time (subframes) and frequency resources where access requests are allowed. These parameters are included in the master information block (MIB) and in the system information blocks (SIBs): *PRACH-ConfigSIB,* SIB1, and SIB2. Some of the most relevant parameters included in these SIBs for the RA are listed in Table 1 and will be described where appropriate.

The CBRA procedure is used for initial access, e.g. when a UE attempts to connect for the first time to the 5G network or after a sleep period. This is the main mode used by MTC UEs, as these only maintain the connection with the gNB long enough to perform one or few data transmissions. In contrast, the CFRA procedure is used when major modifications are made to ongoing connections, such as connection re-establishment, handover, and re-synchronization (3GPP TS 38.300 V15.3.1 2018, Section 9.2.6).

**Table 1** Relevant system configuration parameters that must be acquired before the beginning of the RA procedure and the SIBs where these are included (3GPP TS 36.331 V15.0.0 2018).

| Parameter | In SIB | |
|---|---|---|
| | 5G and LTE-M | NB-IoT |
| Period between SI transmissions | 1 | 1-NB |
| Available set of (N)PRACH resources to transmit preambles within the radio frames | PRACH-Config | NPRACH-Config |
| Maximum number of preamble transmissions per access attempt | 2 | NPRACH-Config |
| Length of the RAR window | RACH-Config-Common | RACH-Config-Common-NB |
| Number of preambles available for the CBRA $K$ | RACH-Config-Common | NPRACH-Config |
| ACB parameters: ac-BarringFactor $p_{\text{acb}}$ and ac-BarringTime $t_{\text{acb}}$ | 2 | — |
| EAB and AB parameters: barring bitmap | 14 | 14-NB |

## Contention-Based Random Access Procedure

The CBRA procedure comprises a four-message handshake between the UEs and the BS (denoted as gNB in 5G). Figure 1 summarizes the message exchange during the CBRA procedure, whose steps are described below (3GPP TS 36.321 V15.2.0 2018, Section 5.1).

### Preamble Transmission (MSG1)

Preambles are orthogonal resources used to perform the RA request (MSG1). They are selected uniformly at random by accessing UEs and can only be transmitted in random access opportunities (RAOs), which correspond to an occurrence of the PRACH. Specifically, RAOs are subframes where RA requests are allowed. The period between RAOs is indicated by the value of the PRACH-ConfigIndex parameter; examples of typical values are illustrated in Figure 2 (3GPP TS 36.211 V15.4.0 2018; Laya et al. 2016).

Preambles in 4G and 5G are generated using Zadoff-Chu sequences. These are constant amplitude zero autocorrelation (CAZAC) sequences; their autocorrelation function is periodic, which allows for accurate preamble detection and timing. Furthermore, Zadoff-Chu sequences exhibit good cross-correlation properties. Hence, preambles can

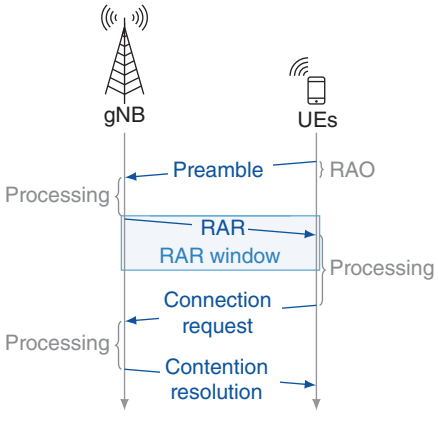


**Figure 1** Four-message exchange that occurs in the CBRA procedure of 5G.

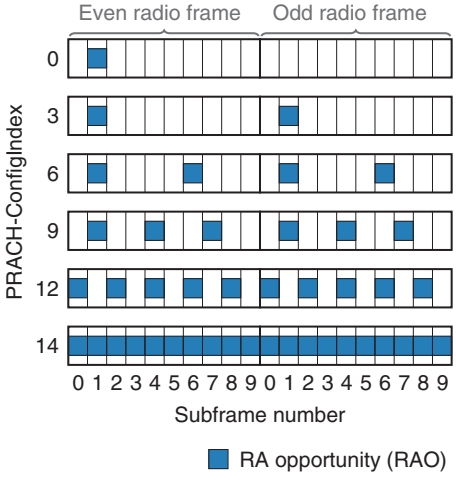


**Figure 2** Period between RAOs according to the PRACH-ConfigIndex parameter (3GPP TS 36.211 V15.4.0 2018; Laya et al. 2016).

be rapidly decoded at the gNB. However, Zadoff-Chu sequences are difficult to generate in real time and require large amounts of memory for their storage (3GPP TS 36.211 V15.4.0 2018); thus, the number of available orthogonal preambles is limited and is signaled by the gNB through the RACH-Config Common message (3GPP TS 36.331 V15.0.0 2018).

The orthogonality of preambles implies that multiple UEs can access the gNB in the same RAO if they select different preambles. Therefore, the simplified channel model for the PRACH (at a given RAO) corresponds to that of multichannel slotted ALOHA, given by (5), where $K$ corresponds to the number of available preambles. In this model, a collision occurs when multiple UEs transmit the same preamble at the same RAO, but it may be undetected by the gNB, as explained later.

The UEs determine the power for the first preamble transmission according to the strength of the reference signals provided by the gNB. To reduce the probability of subsequent preamble transmission failures due to channel errors, UEs perform power ramping, i.e. increase the transmission power, after each failed access attempt.

#### RA Response (MSG2)

The gNB attempts to decode the transmitted preambles and sends up to one random access response (RAR) message per subframe through the Physical Downlink Control Channel (PDCCH). The RAR includes, among other information, uplink grants for the transmission of the next message of the RA procedure in predefined time–frequency resources. A single RAR may contain several uplink grants, each of them associated with a successfully decoded preamble. Therefore, no uplink grant will be sent in response to the collided preambles if the collision is detected.

The PDCCH resources are limited, and there is a maximum number of uplink grants that can be sent per RAR. After the preamble transmission, a UE waits for a predefined number of subframes to receive the uplink grant. This period is known as the RAR window. The number of available uplink grants per RAR window is the product of the length of the RAR window and the number of uplink grants that can be sent per subframe.

#### Connection Request (MSG3) and Contention Resolution (MSG4)

After receiving the uplink grant, the UEs adjust their uplink transmission time according to the received time alignment and schedule the transmission of MSG3 in the dedicated resources. MSG3 contains a unique identifier for each UE. Finally, the gNB transmits a MSG4 in response to each received MSG3. MSG3 and MSG4 transmissions are protected with hybrid automatic repeat request (HARQ). If MSG3 fails, the gNB will not send the MSG4. In this case, the HARQ timer of the UE will expire; at this point, the UE schedules a new MSG3 retransmission. A UE stops retransmitting MSG3 if it does not receive an MSG4 within a predefined time window, or if the maximum number of transmission attempts is reached, resulting in a failed access attempt.

The channel model defined in (5) implies that the gNB decodes the preambles received with sufficient power from exactly one UE in each RAO, while a collision occurs when multiple UEs transmit the same preamble simultaneously. Whether the collision will be detected depends on the channel state; thus, the channel coefficient $h_n$ cannot be neglected. Furthermore, depending on the difference on the power-delay profile, the preamble transmissions of multiple UEs may be additive or destructive. That is, multiple

transmissions of the same preamble can be either interpreted as a single transmission with high power or these may simply not be decoded.

Specifically, a failed access attempt can happen for the following reasons:

- No uplink grant is received within the RAR window:
  - Lost preamble transmission due to low SNR (i.e. transmission power is too low).
  - Insufficient resources in the PDCCH to schedule an uplink grant for a successfully decoded preamble.

  Preamble collision:

  Multiple transmissions of the same preamble interfere with each other.

  Multiple transmissions of the same preamble are received in such a way that the collision is detected by the gNB.
- The UE declares a failure at the contention resolution:
  - Lost MSG3 and/or MSG4 transmissions due to low SNR.
  - Preamble collision: Multiple transmissions of the same preamble present a power delay profile such that the collision cannot be detected. Here the gNB may transmit an uplink grant in response to the decoded preamble. Then, multiple UEs will receive the same uplink grant and transmit their MSG3s in the same predefined resources, leading to detectable collisions and, thus, no MSG4 is transmitted.

Therefore, the probability that a single access attempt following the CBRA procedure defined for 5G is successful can be calculated as

$$\Pr[\mathcal{S}_{\text{CBRA}}] = \Pr[\mathcal{S}_1]\Pr[\mathcal{S}_2](1-(1-\Pr[\mathcal{S}_3]\ \Pr[\mathcal{S}_4])^M) \tag{6}$$

where $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$, and $\mathcal{S}_4$ denote the event of a success at each step of the CBRA and $M$ is the maximum number of allowed MSG3/MSG4 transmissions.

There is a maximum number of allowed preamble transmissions, broadcasted by the gNB through the SIB2. UEs terminate the RA procedure if this limit is reached without success. Conversely, whenever an access attempt fails and, if the maximum number of preamble transmissions has not been reached, the UE first backs off for a random time period and then randomly selects and transmits a new preamble at the next RAO. The backoff time is selected uniformly randomly between 0 and the value of the backoff indicator that is transmitted by the gNB within the RAR messages (3GPP TS 36.321 V15.2.0 2018, Section 6.2.2).

### Contention-Free Random Access

The main difference between the CBRA and CFRA in 5G is that the latter incorporates the transmission of a MSG0 before MSG1. Here MSG0 is transmitted by the gNB to assign a preamble to a specific UE. This preamble is selected from a pool of reserved preambles, which cannot be selected by other UEs, and thus no preamble collisions can occur.

The CFRA continues with the preamble transmission by the UE and concludes with the RAR from the gNB, assigning an uplink grant to the UE. Note that the CFRA has an impact on the CBRA because (i) the number of preambles reserved for the CFRA reduces the number of available preambles for CBRA and (ii) uplink grants are shared, where UEs in the CFRA are prioritized. That is, uplink grants are first sent to UEs with pre-assigned preambles.

### MTC Specific Implementations: NB-IoT and LTE-M

NB-IoT and LTE-M are mMTC-targeted implementations by the 3GPP, included in Release 13 in 2016 (Wang et al. 2017). The enhancements in LTE-M and NB-IoT with respect to legacy LTE mainly focus on the physical layer, aiming to provide greater power efficiency, lower bandwidth utilization, and enhanced coverage at a reduced hardware cost.

To achieve these goals, the bandwidth-reduced low complexity (BL) UE category, also known as category M1, was defined for LTE-M, and the NB-IoT UE category, also known as category M2, was defined for NB-IoT. However, there are minor to no enhancements in the upper layers of LTE-M and NB-IoT when compared to legacy LTE.

In particular, the GBRA defined for legacy LTE remains the same for both BL and NB-IoT UEs, except for the following changes in the physical layer (3GPP TS 36.211 V15.4.0 2018, Section 5.7)

1) Preambles in NB-IoT are single-tone frequency hopping patterns instead of sequences.
2) Preambles may be repeated, one after the other, depending on the channel quality indicator (CQI) of a specific UE.

In particular, the total bandwidth for NB-IoT is 180 kHz (hence "narrowband"), divided into 48 available subcarriers (tones). A preamble comprises four symbol groups, each comprised of a cyclic prefix and 5 single-tone symbols. The duration of the cyclic prefix is either 66.7 or 266.67 μs; the latter is also the duration of each symbol (3GPP TS 36.211 V15.4.0 2018, Section 10.1.6). The first single-tone symbol is selected randomly and subsequent tones depend on the first one. Consequently, a preamble in NB-IoT is defined by its initial tone and there are up to 48 available preambles in NB-IoT.

Three different coverage enhancement (CE) levels are defined: from 0 to 2. Each UE is assigned to a CE level depending on the quality of its wireless link to the gNB, which is CE 0 for the best and CE 2 for the worst one. For example, UEs that repeatedly fail the CBRA procedure increase their CE level. Hence, UEs in a higher CE level perform more preamble repetitions than those in lower CE levels to increase the probability of delivering the preamble to the gNB (Harwahyu et al. 2017; Jiang et al. 2018).

Preamble repetitions are performed one after the other. In LTE-M, BL UEs transmit the same preamble until the specified number of repetitions is reached. In NB-IoT, the pseudo-random frequency hopping patterns used for preamble construction are directly extended for preamble repetitions. Specific preambles are assigned to each CE level, so a given preamble can only be transmitted by UEs in the same CE level. As a consequence, the number of available preambles does not increase with the number of preamble repetitions neither in NB-IoT nor in LTE-M.

## Random Access in mMTC Applications

In this section, we evaluate the efficiency of the CBRA defined for 5G (which is identical to legacy LTE). Let $K$ be the number of available preambles and $A = \sum_{n=1}^{N} a_n$ be the number of active UEs at a given RAO. Next, let $S$ be the random variable (RV) that

defines the number of preambles selected by a single UE at a given RAO. $S$ depends only on the number of preambles $K$ and $A$.

Next, let $T$ be the throughput of the first step of the RA procedure, defined as the expected value of RV $S$ for a given value of $K$ and $A$. That is,

$$T(K,A) = \mathbb{E}[S|K,A] = \sum_{s=1}^{K} s \Pr[S = s|K,A] \tag{7}$$

The probability mass function (pmf) of $S$ can be precisely calculated via a recursive approach that provides the least computational complexity (Tello-Oquendo et al. 2019); numerous other methods have been also proposed for this task (Wei et al. 2015). In this article, we derive $T$ via the probability that exactly UE out of $A$ transmits a given preamble:

$$T_k(K,A) = \binom{A}{1} \frac{1}{K}\left(1 - \frac{1}{K}\right)^{A-1} \approx \frac{A}{K} e^{-\frac{A}{K}} \tag{8}$$

Since there is a total of $K$ preambles, we have

$$T(K,A) = K\ T_k(K,A) = A\left(1 - \frac{1}{K}\right)^{A-1} \tag{9}$$

It is easy to verify that the maximum throughput is achieved when $A = K$, that is,

$$T_{\max}(K) = \max_{A} T(K,A) = T(K,K) = K\left(1 - \frac{1}{K}\right)^{K-1} > \frac{K}{e} \tag{10}$$

Nevertheless, MSG2 additionally restricts the throughput of the CBRA, through the number of uplink grants that can be transmitted per RAR window, hereafter denoted as $G$. Define the RV $S_g$ as the number of uplink grants assigned in response to preambles transmitted at a given RAO. The maximum throughput of the CBRA is (Leyva-Mayorga et al. 2019)

$$T_{\max}(K,G) = \mathbb{E}[S_g|\mathrm{K},\mathrm{G}] \tag{11}$$

Naturally, the critical point for $T_{\max}(K,G)$ is also $K$.

As observed in the following section, the main objective of access control in 5G is to set the number of active UEs after the access control (i.e. the number of UEs that perform the CBRA) to the optimal value of $A$ for any $K$, denoted as $A^* = \min\{A,K\}$. Doing so maximizes resource utilization while minimizing the RA latency.

However, achieving even a near-optimal performance is a complicated task because the following characteristics are inherent to the CBRA:

1) The total number of UEs, within the area of coverage of a specific cell, may not be known.
2) The number of active UEs per subframe $A$ is not known.
3) Active users lack a coordination mechanism among themselves.
4) There can be numerous reasons for a failed access attempt (see the section titled "Contention-Based Random Access Procedure").

The cases with $A > K$ are highly problematic, since congestion builds up and performance severely deteriorates, as observed in a technical report (3rd Generation Partnership Project 2011). In the latter report, the number of UEs within the coverage area of a BS was calculated, based on a specific application of smart electric metering in central

**Table 2** Characteristics of the traffic models proposed by the 3GPP for the evaluation of the RA procedure under mMTC applications (3rd Generation Partnership Project 2011).

| Parameter | Traffic model 1 | Traffic model 2 |
|---|---|---|
| Number of MTC UEs | $\{1, 3, 5, 10, 30\} \cdot 10^3$ | |
| Distribution period (s) | 60 | 10 |
| Distribution over the period | Uniform | Beta(3, 4) |

and urban areas of London. The 3GPP observed that more than 30 000 UEs may be connected to a single macro BS and developed two traffic models to evaluate the performance of the RA procedure of LTE, denoted by traffic models 1 and 2.

The traffic model 1 corresponds to a weakly synchronized mMTC scenario, in which the access attempts of UEs are uniformly distributed over a relatively long period of 60 seconds. On the other hand, the traffic model 2 represents a highly synchronized scenario, in which the access attempts of UEs follow a Beta(3, 4) distribution over a relatively short period of 10 seconds. Table 2 presents the characteristics of these models.

If the number of UE arrivals scheduled within a period is large, the peak of arrivals per subframe can easily exceed $T_{\max}(K, G)$ in the traffic model 2. To illustrate this, let $X_c$ be the continuous RV with Beta($\alpha, \beta$) distribution that defines the time elapsed between the start of the period and the arrival of a specific UE, where the support for $X_c$ is the period $\tau = [0, \tau_{\max}]$ s. The distribution of $X_c$ is

$$f_{X_c}(\tau) = \frac{\tau^{\alpha-1}(\tau_{\max} - \tau)^{\beta-1}}{B(\alpha, \beta)\tau_{\max}^{\alpha-\beta-1}} \tag{12}$$

where $B(\alpha, \beta) = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$ is the Beta function.

Next, let $X$ be the discrete RV with Beta(3, 4) distribution, which defines the number of subframes between the beginning of the period and the arrival of a specific UE, and $\Delta x$ the period between RAOs. The pmf of $X$, whose support is $x = [0, x_{\max} = \lfloor \tau_{\max}/\Delta x \rfloor]$, can be obtained by integrating the pmf of $X_c$ between the time of the beginning of subframes $x$ and $(x + 1)$. Next, let $\{Y_x\}_{x \in \{0,1,\ldots,x_{\max}\}}$ be the stochastic process that defines the number of UE arrivals that occur at the $x$th RAO. That is, the number of UEs that will initiate the CBRA procedure at the $x$th RAO. For the suggested values of $\alpha = 3$ and $\beta = 4$, we have

$$\mathbb{E}[Y_x] = Np_X(x) = \frac{60Nx^2(x_{\max} - x)^3}{x_{\max}^6 - x_{\max}^2} \tag{13}$$

It is easily obtained that the maximum of (13) occurs at $x^* = 2x_{\max}/5$, where

$$\mathbb{E}[Y_{x^*}] = \frac{1296Nx_{\max}^3}{625(x_{\max}^4 - 1)} \tag{14}$$

Several studies have shown that if $N = 30\,000$ UE arrivals occur according to the traffic model 2 under a typical RACH configuration, only around 31% of the UEs will successfully complete the CBRA procedure (3rd Generation Partnership Project 2011; Leyva-Mayorga et al. 2017; Wei et al. 2015). Therefore, access control mechanisms are needed to support mMTC applications.

## Access Control Mechanisms Defined for 5G Networks

Most of the research in access control methods for the CBRA has focused on the highly synchronized access, described by the traffic model 2, as this is the scenario that is the most prone to severe congestion episodes. However, congestion can also occur if the access requests are weakly or even non-synchronized. Therefore, intensive research and standardization work has been done to mitigate congestion and maximize resource utilization, where the standardization efforts have focused on barring schemes (3GPP TS 22.011 V16.4.0 2018, Section 4; 3GPP TS 36.331 V15.0.0 2018, Sections 5.3.3.11-14).

The remainder of this section provides a summary of the barring schemes defined for 5G. Afterward, in the section titled "Improvements to Existing Protocols" presents some improvements to these schemes for efficient mMTC access.

### Access Class Barring (ACB)

The access class barring (ACB) scheme provides a mechanism to delay the access requests of UEs, in order to reduce the number of simultaneous access attempts per RAO (3GPP TS 22.011 V16.4.0 2018, Section 4.3.1) and redistribute them over time. In ACB, gNB instructs the UEs to randomly delay the beginning of their RA attempt according to the barring parameters, which are the barring factor $p_{\text{acb}}$ and mean barring time $t_{\text{acb}}$, transmitted through the SIB2 (3GPP TS 36.331 V15.0.0 2018, pp. 387–392).

All UEs are members of one randomly allocated access classes (ACs) from 0 to 9. In addition, specific high priority UEs may be members of one or more out of five special ACs (11–15) (3GPP TS 22.011 V16.4.0 2018). Barring parameters apply to all ACs 0–9 and, if so indicated, to one or more of the ACs 11–15. In other words, even though the numbering suggests a distinction between ACs, only special ACs are treated differently, whereas normal ACs are treated equally.

The UEs must continuously acquire the SIB2 and update the barring parameters accordingly. SIB2 is transmitted once every si-Periodicity frame as indicated in the SIB1. The UEs subject to the ACB scheme perform a barring check immediately before initiating the CBRA with the transmission of the first preamble; see Figure 3 (3GPP TS 36.331 V15.0.0 2018, Section 5.3). The UEs that pass a barring check are no longer subject to the ACB scheme and initiate the CBRA as described in the section titled "Contention-Based Random Access Procedure". On the other hand, the UEs that fail the barring check must wait for a random period and perform a new barring check afterward (3GPP TS 22.011 V16.4.0 2018).

It should be noted that NB-IoT UEs are not subject to the ACB scheme, but to the access barring (AB) scheme, which is described in the section titled "Extended Access Barring (EAB)", along with the extended access barring (EAB) scheme.

The Application-specific Congestion control for Data Communication (ACDC) is another access control scheme that aims to provide application specific barring. Here, the network operator defines a set of categories, ranked depending on the desired level of restriction: the highest category will be the least restricted and vice versa. Apart from this, the operation of the ACDC scheme is similar to that of the ACB scheme, and the former can be seen as a generalization of the latter (3GPP TS 22.011 V16.4.0 2018, Section 4.3.5).

If the barring parameters are set correctly, the ACB may be effective to relieve sporadic and relatively short periods of congestion (in the order of a few seconds), even

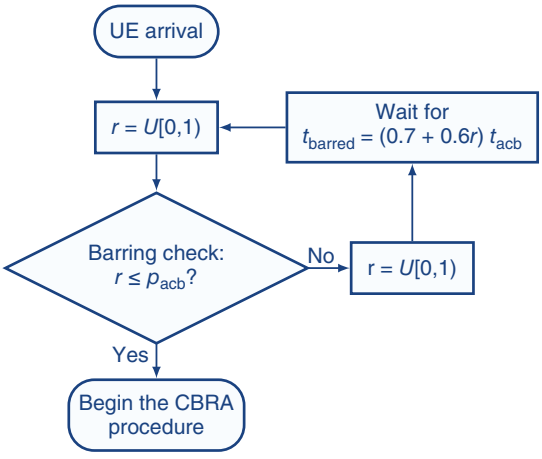


**Figure 3** Description of the ACB scheme. The UE is only allowed to begin the RA procedure once it performs a successful barring check (if $r \leq p_{acb}$).

if the barring parameters are selected beforehand, based on a forecast of the number and distribution of UE arrivals, and remain fixed throughout the network operation (Tello-Oquendo et al. 2018). However, such approach induces an unnecessary delay to the access of UEs that occur during periods when just a few arrivals per RAO occur, that is, when $A \ll K$. Therefore, barring parameters should be selected according to the number of active UEs per RAO. However, 5G neither has embedded mechanisms to automatically calculate the barring parameters, nor has the 3GPP provided any recommendations for their selection. Consequently, methods to adequately select values for the barring parameters were proposed in the literature (Duan et al. 2016; Leyva-Mayorga et al. 2019; Lin et al. 2014; Tavana et al. 2018), and their efficiency with respect to other techniques was studied (De Andrade et al. 2017). Hereafter, we refer to these methods as Adaptive Barring Configuration (ABC) schemes; we describe and classify them in the section titled "Adaptive Barring Configuration (ABC) Schemes".

#### Extended Access Barring (EAB)

The EAB scheme provides a deterministic mechanism to prevent certain ACs from attempting access for a predefined period and applies to both normal and BL UEs. EAB parameters are broadcast in the SIB14, conveying also the set of barred ACs (3GPP TS 36.331 V15.0.0 2018, Section 5.2.1.7).

When the EAB configuration parameters change, not all UEs acquire the new parameters from the next SIB14 transmission. Instead, the UEs are randomly distributed among the SIB14 transmissions over a predefined period. The purpose of this is to reduce the number of UEs that will perform a simultaneous access attempt after the barring of an AC is lifted (i.e. changed from barred to not barred). However, lifting an AC still induces synchronization among UEs belonging to this AC. These UEs will initiate their access procedure in bursts, which causes numerous preamble collisions during the first RAOs and deteriorates the overall network performance (Cheng et al. 2015).

The AB is the barring mechanism specifically defined for NB-IoT. The AB parameters are included in the SIB14-NB (SIB14 for NB-IoT) and, as in EAB, a bitmap is used to indicate the barred ACs. In addition, the AB scheme also indicates a threshold for the reference signal received power (RSRP). If the RSRP is lower than the specified threshold, the access of the UE is barred. By doing so, the access of UEs with particularly poor

wireless conditions can be restricted, and the RA resources are made available only for UEs with relatively low error probability of transmission.

## Improvements to Existing Protocols

Maximization of resource utilization requires continuous adaptation of the barring parameters to the number of contending UEs; we proceed by formally defining the characteristics of an optimal access control scheme. Recall that $A$ denotes the number of active UEs when no access control is applied and $A = K$ maximizes the throughput of the CBRA. Next, we formally define $A_c \leq A$ as the number of active UEs at a RAO after access control. Recall that $A^* = \min\{A, K\}$ is the optimal value of $A$. The optimal access control scheme fulfills $A_c = A^*$. This not only prevents congestion, when a high number of access attempts per subframe occur by barring users from access, but also allows the access of all active UEs in normal operating conditions, where no congestion occurs.

### Resource Separation and Prioritization

One of the main concerns in mMTC is how to guarantee that the latency requirements of diverse coexisting applications are met. Resource separation and prioritization is a straightforward solution to this problem. In turn, an intuitive form of resource separation is preamble separation, where a subset of preambles is reserved for latency-sensitive MTC applications or for H2H services (Lin et al. 2014) and the rest of the preambles are used for applications with lower QoS requirements. While high priority UEs may benefit from these schemes, it is evident that congestion cannot be prevented simply by separating preambles as the achievable throughput $T_{\max}(K, G)$ is not affected.

### Adaptive Barring Configuration (ABC) Schemes

ABC schemes attempt to fine-tune the barring parameters based on the outcome of the CBRA and can be classified in two main groups. The first group attempts to estimate the number of active UEs at each RAO; we refer to these as *estimation-based schemes*. The second group merely calculates the barring parameters from the number of successful accesses and no estimation is performed; we refer to these as *throughput-based schemes*. Most of the research in this domain has focused on the ACB scheme; nevertheless, the principles described in the following are also applicable to the EAB scheme.

To implement an estimation-based scheme, one or several of the following parameters must be known or estimated firstly:

1) The number of preambles transmitted by a single UE (i.e. successful preambles);
2) The number of preambles transmitted by multiple UEs (i.e. collided preambles);
3) The number of preambles not selected by any UE (i.e. idle preambles);
4) The number of UEs previously registered in the cell.

By using some of these parameters, the number of active UEs can be estimated. Then, either the barring rate $p_{\text{acb}}$ (in ACB) or the number of barred ACs (in EAB) can, in theory, be selected to achieve $A_c = A^*$, which leads to an optimal performance. Throughput-based schemes use the number of successful preambles $S$ as the single

load indicator and attempt to maintain this number below, but as close as possible to $T_{\max}(K, G)$.

Most of the literature on ABC schemes focuses on estimation-based schemes, proposing several techniques to enhance their precision for the ACB scheme (Duan et al. 2016; Tavana et al. 2018). Although such techniques provide greater level of insight and a finer control to the ACB parameters than the throughput-based schemes, they are difficult to implement since they are generally based on two assumptions that do not match the actual operation of the ACB scheme defined in the 3GPP standards:

1) The barring parameters are updated on a RAO-by-RAO basis.
2) Every UE is subject to the ACB scheme.

In 3GPP standards, barring parameters are updated once every si-Periodicity frame and the minimum possible setting for this parameter is 8. Hence, barring parameters can be updated up to once every

$$t_{\text{update}} = \frac{\text{si-Periodicity} \times 10}{\text{RACH-Periodicity}} \text{ [RAOs]} \tag{15}$$

Under a typical network configuration (3rd Generation Partnership Project 2011), the period between RAOs RACH-Periodicity is set to five subframes, which yields $t_{\text{update}} = {}^{80}/_{5} = 16$ RAOs. On the other hand, estimation schemes generally assume $t_{\text{update}} = 1$, which leads to the theoretical upper bound in performance of the scheme but fails to provide meaningful information on its performance in a practical implementation.

A simple and effective solution to extend estimation-based schemes to practical values of $t_{\text{update}}$ is to use the average of $A$ over the update period to select $p_{\text{acb}}$ (Leyva-Mayorga et al. 2019). That is,

$$p_{\text{acb}}^{*} = \min\left\{1, \frac{A^{*}}{\overline{A}}\right\} \tag{16}$$

where

$$\overline{A} = \frac{1}{t_{\text{update}}} \sum_{i=1}^{t_{\text{update}}} A_i \tag{17}$$

and $A_i$ denotes the number of active UEs at the $i$th RAO of the update period.

The second assumption, that every UE is subject to the ACB scheme, turns out to be more problematic. In 3GPP standards, the ACB scheme only affects the UEs that have not yet transmitted their first preamble, but not the UEs that are reattempting preamble transmission. To illustrate this problem, let $A^{(j)}$ be the number of contending UEs that are about to perform their $j$th preamble transmission, where $A = \sum_j A^{(j)}$, $j \in \{1, 2, \ldots, j_{\max}\}$, and $j_{\max}$ is the maximum number of preamble transmissions. At each RAO, the ACB scheme only affects $A^{(1)}$. Methods for an accurate estimation of $A$ at each RAO that use no information of its past values have been reported in the literature (Duan et al. 2016; Tavana et al. 2018; Tello-Oquendo et al. 2019). However, estimating both $A$ and $A^{(1)}$ requires the sequential estimation over a sufficiently long number of consecutive RAOs.

Finally, it is generally assumed in estimation-based schemes that no preamble transmissions are lost due to wireless channel errors. That is, the simplified channel model

in (5) is used. If wireless channel errors can occur, the complexity of the estimation increases.

The throughput-based schemes are much simpler, relying on the use of the number of successful accesses to directly calculate a load indicator, which gives a rough approximation of the state of the system. Hence, the operation of throughput-based schemes can be summarized as follows:

1) Calculate the maximum throughput of the RA procedure $T_{\max}(K, G)$.
2) Store the number of successful accesses $S$ at each RAO.
3) Calculate a load indicator $L$.
4) Calculate $p_{\text{acb}} = f(L)$, where $f$ is a decreasing function.

Since barring parameters are updated at each SIB2 transmission, which occurs every $t_{\text{update}}$ RAOs, the ratio of utilized to available resources within this period can be used as the load indicator:

$$L = \frac{1}{t_{\text{update}} T_{\max}(K, G)} \sum_{i=1}^{t_{\text{update}}} S_i \tag{18}$$

Such load indicator is highly simplistic and can be further refined to stabilize the barring parameters due to the following issues:

1) The randomness of the number of active UEs $A$, of the preamble selection, and of the wireless channel may induce relatively high variations in $S$. This in turn may lead to highly variable barring parameters and, hence, to a suboptimal access control.
2) Using the values of $S$ just from a single update period to calculate the load indicator may be inadequate. For instance, two contradictory reasons exist for the load indicator in (18) to decrease from one period to the next one:
   a) The barring parameters were selected correctly and $A$ decreases.
   b) The barring parameters were not selected properly and the values of $S$ during the period decrease due to congestion. This occurs when $A_c \gg K$.

To illustrate the latter issue, Figure 4 shows the expected number of successful accesses $T(K, G, A) = \mathbb{E}[S|K, G, A]$ for $G = 15$ and for two characteristic values of $K = \{24, 54\}$; the former is a possible setting for 5G, LTE-M, and NB-IoT, whereas the latter is widely

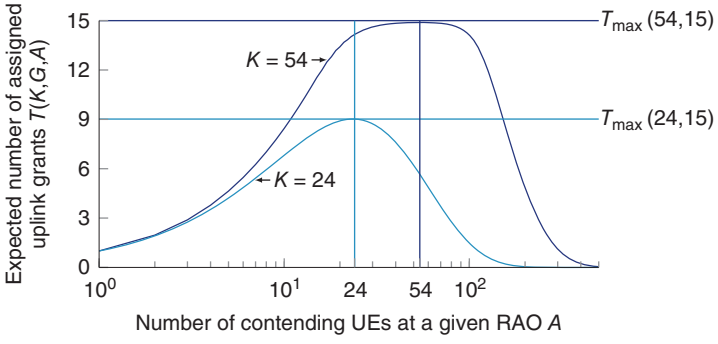


**Figure 4** Expected number of successful accesses as a function of the number of active UEs $T(K, G, A) = E[S|\, K, G, A]$, given $K = \{24, 54\}$ and $G = 15$.

used in the LTE-A literature. Clearly, two different values of $A$ lead to same $T(K, G, A)$, as elaborated in (Lin et al. 2014).

## Emerging Technologies and Conclusions

The current method for initial access to the 5G BSs is inefficient to handle mMTC and URLLC MTC applications. Furthermore, mMTC-specific implementations merely provide benefits in terms of power consumption, spectral efficiency, and coverage. While numerous enhancements to barring schemes included in the 3GPP standards have been proposed, there are still some issues to be solved and major modifications or enhancements are needed in the second phase of standardization to fulfill the bold promises of 5G technology.

A promising solution for short packet transmission in mMTC and URLLC applications is to adopt a GFRA protocol (METIS-II 2017; Popovski et al. 2018) and develop the mechanisms to allocate resources to both the existing GBRA in 5G and the new GFRA, so the UEs can select which to use. Multipacket reception techniques can be used to complement the GFRA. This would provide major benefits in terms of complexity, latency, energy, and scalability to MTC applications. At the present time, there is no consensus on the preferred GFRA protocol nor on how to incorporate such a protocol to the current 5G architecture/standards.

Another promising approach to support mMTC is to enhance the standard CBRA with cooperative RA, where some of the UEs are selected as group heads. The group heads use short-range links to aggregate access requests from neighboring UEs and are the only ones that perform the CBRA procedure. As a consequence, the number of direct access requests to the gNB decreases significantly without the need to deploy additional infrastructure (5G Infrastructure Public Private Partnership 2017). Cooperative RA is in line with one of the promises of 5G: to provide full integration with 4G and short-range technologies. However, the overhead of cooperative RA is the increased energy consumption at the group heads and the main challenge relies on identifying and providing a proper incentive to get UEs to cooperate.

Other approaches include using a distributed queuing algorithm to achieve a much better scalability than traditional ALOHA (Laya et al. 2016), as well as combining sequences of preambles into user signatures to increase the contention space and, potentially, to convey additional information to the BS (Pratas et al. 2012, 2017).

## Acknowledgments

The work of I. Leyva-Mayorga, C. Stefanovic, and P. Popovski has been partly supported by the European Research Council (ERC) under the European Union Horizon 2020 research and innovation program (ERC Consolidator Grant Nr. 648382 WILLOW), by the Horizon 2020 project ONE5G (ICT- 760809), and by the Danish Council for Independent Research (Grant Nr. 8022-00284B SEMIOTIC). The work of V. Pla and J. Martinez-Bauset was supported by Grant PGC2018-094151-B-I00 (MCIU/AEI/FEDER, UE).

## Related Articles

Hybrid ARQ Schemes
Orthogonal Multiple Access
Random Access Protocols
Random Access for Cellular Systems

## References

3GPP TS 22.011 V16.4.0 (2018). *Service Accessibility*. Sophia Antipolis Valbonne, France: 3rd Generation Partnership Project.

3GPP TS 22.368 V14.0.1 (2017). *Service Requirements for Machine-Type Communications*. Sophia Antipolis Valbonne, France: 3rd Generation Partnership Project.

3GPP TS 36.211 V15.4.0 (2018). *Physical Channels and Modulation*. Sophia Antipolis Valbonne, France: 3rd Generation Partnership Project.

3GPP TS 36.321 V15.2.0 (2018). *Medium Access Control (MAC) Protocol Specification*. Sophia Antipolis Cedex, France: 3rd Generation Partnership Project.

3GPP TS 36.331 V15.0.0 (2018). *Radio Resource Control (RRC); Protocol Specification*. Sophia Antipolis Valbonne, France: 3rd Generation Partnership Project.

3GPP TS 36.912 V15.0.0 (2018). *Feasibility Study for Further for E-UTRA*. Sophia Antipolis Valbonne, France: 3rd Generation Partnership Project.

3GPP TS 38.300 V15.3.1 (2018). *NR; NR and NG-RAN Overall Description; Stage 2*. Sophia Antipolis Valbonne, France: 3rd Generation Partnership Project.

3rd Generation Partnership Project (2011). Study on RAN improvements for machine-type communications. 3GPP TR 37.868.

3rd Generation Partnership Project (2017). 5G; Study on scenarios and requirements for next generation access technologies. 3GPP TR 38.913 V14.2.0.

5G Infrastructure Public Private Partnership (2017). Final Overall 5G RAN Design. 5G-PPP METIS-II Report 2.4.

Cheng, R.-G., Chen, J., Chen, D.-W. et al. (2015). Modeling and analysis of an extended access barring algorithm for machine-type communications in LTE-A networks. *IEEE Transactions on Wireless Communications* 14 (6): 2956–2968. doi: 10.1109/TWC.2015.2398858.

Cisco (2017). Cisco visual networking index (VNI): forecast and trends, 2017–2022. https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html (accessed 04 April 2019).

De Andrade, T.P.C., Astudillo, C.A., Sekijima, L.R. et al. (2017). The random access procedure in Long Term Evolution networks for the Internet of Things. *IEEE Communications Magazine* 55 (3): 124–131. doi: 10.1109/MCOM.2017.1600555CM.

Duan, S., Shah-Mansouri, V., Wang, Z. et al. (2016). D-ACB: adaptive congestion control algorithm for bursty M2M traffic in LTE networks. *IEEE Transactions on Vehicular Technology* 65 (12): 9847–9861. doi: 10.1109/TVT.2016.2527601.

Harwahyu, R., Cheng, R.-G., and Wei, C.-H. (2017). Investigating the performance of the random access channel in NB-IoT. Proceedings of the IEEE 86th Vehicular Technology Conference (VTC-Fall), 1–5. doi: 10.1109/VTCFall.2017.8288195.

Jiang, N., Deng, Y., Condoluci, M. et al. (2018). RACH preamble repetition in NB-IoT network. *IEEE Communications Letters* 22 (6): 1244–1247. doi: 10.1109/LCOMM.2018.2793274.

Laya, A., Alonso, L., and Alonso-Zarate, J. (2014). Is the random access channel of LTE and LTE-A suitable for M2M communications? A survey of alternatives. *IEEE Communications Surveys & Tutorials* 16 (1): 4–16. doi: 10.1109/SURV.2013.111313.00244.

Laya, A., Kalalas, C., Vazquez-Gallego, F. et al. (2016). Goodbye, ALOHA! *IEEE Access* 4: 2029–2044. doi: 10.1109/ACCESS.2016.2557758.

Leyva-Mayorga, I., Tello-Oquendo, L., Pla, V. et al. (2017). On the accurate performance evaluation of the LTE-A random access procedure and the access class barring scheme. *IEEE Transactions on Wireless Communications* 16 (12): 7785–7799. doi: 10.1109/TWC.2017.2753784.

Leyva-Mayorga, I., Rodriguez-Hernandez, M.A., Pla, V. et al. (2019). Adaptive access class barring for efficient mMTC. *Computer Networks* 149: 252–264. doi: 10.1016/j.comnet.2018.12.003.

Lin, T.M., Lee, C.H., Cheng, J.P. et al. (2014). PRADA: Prioritized random access with dynamic access barring for MTC in 3GPP LTE-A networks. *IEEE Transactions on Vehicular Technology* 63 (5): 2467–2472. doi: 10.1109/TVT.2013.2290128.

LoRa Alliance Technical Commitee (2017). *LoRaWAN 1.1 Specification.* Oregon: LoRa Alliance.

METIS-II (2017). Final Overall 5G RAN Design. 5G PPP METIS-II Report 2.4.

Osti, P., Lassila, P., Aalto, A. et al. (2014). Analysis of PDCCH performance for M2M traffic in LTE. *IEEE Transactions on Vehicular Technology* 63 (9): 4357–4371. doi: 10.1109/TVT.2014.2314532.

Popovski, P., Nielsen, J.J., Stefanovic, C. et al. (2018). Wireless access for ultra-reliable low-latency communication: principles and building blocks. *IEEE Network* 32 (2): 16–23. doi: 10.1109/MNET.2018.1700258.

Pratas, N.K., Thomsen, H., Stefanovic, C. et al. (2012). Code-expanded random access for machine-type communications. Proceedings of the IEEE GLOBECOM Workshops, Anaheim, CA (3–7 December 2012). doi: 10.1109/GLOCOMW.2012.6477838.

Pratas, N.K., Pattathil, S., Stefanovic, C. et al. (2017). Massive machine-type communication (MMTC) access with integrated authentication. Proceedings of the IEEE International Conference on Communications, Paris, France (21–25 June 2017). doi: 10.1109/ICC.2017.7997466.

Tavana, M., Rahmati, A., and Shah-Mansouri, V. (2018). Congestion control with adaptive access class barring for LTE M2M overload using Kalman filters. *Computer Networks* 141: 222–233. doi: 10.1016/j.comnet.2018.01.044.

Tello-Oquendo, L., Leyva-Mayorga, I., Pla, V. et al. (2018). Performance analysis and optimal access class barring parameter configuration in LTE-A Networks with massive M2M traffic. *IEEE Transactions on Vehicular Technology* 67 (4): 3505–3520. doi: 10.1109/TVT.2017.2776868.

Tello-Oquendo, L., Pla, V., Leyva-Mayorga, I. et al. (2019). Efficient random access channel evaluation and load estimation in LTE-A with massive MTC. *IEEE Transactions on Vehicular Technology* 68 (2): 1998–2002. doi: 10.1109/TVT.2018.2885333.

Wang, Z. and Wong, V.W.S. (2015). Optimal access class barring for stationary Machine Type Communication devices with timing advance information. *IEEE Transactions on Wireless Communications* 14 (10): 5374–5387. doi: 10.1109/TWC.2015.2437872.

Wang, Y.-P.E., Lin, X., Adhikary, A. et al. (2017). A primer on 3GPP Narrowband Internet of Things. *IEEE Communications Magazine* 55 (3): 117–123. doi: 10.1109/MCOM.2017.1600510CM.

Wei, C.-H., Bianchi, G., and Cheng, R.-G. (2015). Modeling and analysis of random access channels with bursty arrivals in OFDMA wireless networks. *IEEE Transactions on Wireless Communications* 14 (4): 1940–1953. doi: 10.1109/TWC.2014.2377121.

## Further Reading

Hoglund, A., Lin, X., Liberg, O. et al. (2017). Overview of 3GPP release 14 enhanced NB-IoT. *IEEE Network* 31 (6): 16–22. doi: 10.1109/MNET.2017.1700082.

Lin, X., Adhikary, A., and Eric Wang, Y.-P. (2016). Random access preamble design and detection for 3GPP narrowband IoT systems. *IEEE Wireless Communications Letters* 5 (6): 640–643. doi: 10.1109/LWC.2016.2609914.

Mehmood, Y., Görg, C., Muehleisen, M. et al. (2015). Mobile M2M communication architectures, upcoming challenges, applications, and future directions. *EURASIP Journal on Wireless Communications and Networking* 2015 (1): 1–37. doi: 10.1186/s13638-015-0479-y.

Soltanmohammadi, E., Ghavami, K., and Naraghi-Pour, M. (2016). A survey of traffic issues in machine-to-machine communications over LTE. *IEEE Internet of Things Journal* 3 (6): 865–884. doi: 10.1109/JIOT.2016.2533541.

Verma, P.K., Verma, R., Prakash, A. et al. (2016). Machine-to-machine (M2M) communications: a survey. *Journal of Network and Computer Applications* 66: 83–105. doi: 10.1016/j.jnca.2016.02.016.